

ICR Requirement Analysis

Preliminary results

caGrid Project

Oct 13 2004

Types of Sources

- ▶ Data Source
 - Databases, public domain data repositories, sites for literature search
- ▶ Service Source
 - Analysis and data processing programs
- ▶ Storage Source
 - Data caching and on-demand database creation
- ▶ Compute Source
 - Computational sources to execute services

Some Common Services

- ▶ Access Control
- ▶ Metadata definition management
- ▶ Data Management
- ▶ Advertisement and discovery
- ▶ Data caching and on-demand database creation
- ▶ Query Execution
- ▶ Naming and Translation services
 - ID resolution (a gene is called A by one researcher and B by another)
 - translation from vocabulary to vocabulary or from one vocabulary to a common vocabulary.
- ▶ Provenance tracking and management
- ▶ Distributed workflow management

Some comments from ICR group

- ▶ Security
 - Important but separate issue.
 - Important for some data providers, but not others
- ▶ Workflow
 - Not a priority.
- ▶ Performance
 - The faster the better but not an issue.
 - Fast ID resolution was brought up
 - Performance Criteria has not yet been established

Some comments from ICR group

- ▶ Provenance
 - Difficult because you need to define the context.
 - Probably not a attribute level.
- ▶ Prototype
 - It should include Query, Analysis and Annotation.
 - First data retrieval, then ID resolution and then track version.
- ▶ Dynamic data source schema.
- ▶ Integrate public domain data source in the grid.

Software feature list

Preliminary results

	MSKCC	Wistar (L)	Wistar (J)	Washington	Upenn	Georgetown	UNC
Query caBIG resources	X	X	X	X	X	X	X
Update datasources		X					
Advertise data source		X	X	X	X	X	
Advertise anal. Tool		X	X				X
Discover caBIG resources		X	X	X		X	X
Notification of resource update		X	X		X	X	X
Query literature			X	X			
Annotation			X	X			
Analysis tools		X	X	X			X
ID Resolution	X				X		X
Data mapping		X		X			
Metadata (*)	X	X	X	X	X	X	X
WorkFlow analysis tools		X		X			
Quality control		X					
Security				X	X		
Performance							

* Includes version tracking, provenance, CDE



Adopter



Data provider



Tool provider